

Dimension Reduction Analysis of Vowel Signal Data Based on Manifold Learning

JINQING SHEN

Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China,
Shenzhen, CHINA

ZHONGXIAO LI, XIAODONG ZHUANG

Electronics & Information College, Qingdao University, Qingdao, CHINA

Abstract—Data dimension reduction is an important method to overcome dimension disaster and obtain as much valuable information as possible. Speech signal is a kind of non-stationary random signal with high redundancy, and proper dimension reduction methods are needed to extract and analyze the signal features efficiently in speech signal processing. Studies have shown that manifold structure exists in high-dimensional data. Manifold dimension reduction method aiming at discovering the intrinsic geometric structure of data may be more effective in dealing with practical problems. This paper studies a data dimension reduction method based on manifold learning and applies it to the analysis of vowel signals.

Keywords—dimension reduction, manifold learning, LLE, ISOMAP, vowel signal

Received: November 10, 2020. Revised: April 30, 2021. Accepted: May 16, 2021. Published: June 5, 2021.

1. Introduction

Currently, the amount of available information increases exponentially. It is necessary to effectively reduce the dimension of high-dimensional data in order to solve the "dimension disaster" [1], and obtain as much valuable information as possible [2]. Data dimension reduction studies low-dimensional manifolds embedded in high-dimensional space and their different types, so as to find out the essential properties of data set [3,4]. The purpose of dimension reduction technology based on manifold learning is to find out the inherent geometric structure of high-dimensional data in order to solve practical problems. Tenenbaum, etc. replaced Euclidean distance in MDS algorithm with geodesic distance, and proposed an ISOMAP algorithm based on global attributes [5]. Roweis, etc. considered the local information of samples in the process of data dimension reduction, and proposed a nonlinear LLE algorithm [6]. ISOMAP and LLE have greatly promoted the development of nonlinear dimension reduction methods based on manifold learning.

At present, manifold learning method has been applied to feature extraction of sound signals, and the results show that the essential features of stationary sound signals can be discovered by manifold learning [7]. Speech signal is a non-stationary random signal with high redundancy. In speech signal processing, appropriate dimension reduction methods are also needed for efficient feature extraction and analysis [8]. Therefore, this paper mainly studies the data dimension reduction method based on manifold learning and applies it to the analysis of vowel signals.

2. Introduction of Manifold Learning Methods

2.1 Locally Linear Embedding Algorithm (LLE)

The core of LLE algorithm is to assume that the points adjacent to each other in the high-dimensional sample space are usually located on the same manifold, and LLE algorithm can make the nonlinear high-dimensional data keep the original local linear property as much as possible after dimension reduction.

The LLE algorithm can be summarized in three steps [9]:

Step1. For each data point X_i , compute its k -nearest neighbors (based on Euclidean distance or some other appropriate definition of 'nearness').

Step2. Compute weights W_{ij} that best reconstruct each data point X_i , from its neighbors, minimizing the reconstruction error J :

$$J(W) = \sum_{i=1}^m \|X_i - \sum_{j=1}^k W_{ij} X_j\|^2 \quad (1)$$

Step3. Compute the low dimensional embeddings Y_i , best reconstructed by the weights W_{ij} , minimizing the cost function φ :

$$\varphi(Y) = \sum_{i=1}^m \|Y_i - \sum_{j=1}^k W_{ij} Y_j\|^2 \quad (2)$$

In Step 2, the reconstruction error is minimized subject to two constraints: first, that each input is reconstructed only from its nearest neighbors, or $W_{ij} = 0$ if X_i not a neighbor of X_j ; second, that the reconstruction weights for each data point sum to one, or $\sum_j W_{ij} = 1 \forall i$. The optimum weights for each input can be computed efficiently by solving a constrained least squares problem.

2.2 Isometric Mapping Algorithm (ISOMAP)

The main idea of ISOMAP is to replace Euclidean distance in MDS algorithm with geodesic distance between all data points, and try to ensure the global relationship between data and explore the internal geometric structure on the low-dimensional manifold of samples in the process of dimension reduction, so as to better mine the essential characteristics of samples and reveal their internal manifold information more reasonably.

While ISOMAP and LLE have similar aims, ISOMAP is based on a different principle than LLE. In particular, ISOMAP attempts to preserve the global geometric properties

of the manifold while LLE attempts to preserve the local geometric properties of the manifold [7,9].

As with LLE, the ISOMAP algorithm consist of three steps:

Step1. Construct a neighborhood graph - Determine which points are neighbors on the manifold based on distances $d(i,j)$ between pairs of points i, j in the input space (as in step1 of LLE). These neighborhood relations are then represented as a weighted graph over the data points with edges of weight $d(i, j)$ between neighboring points.

Step2. Compute the shortest path between all pairs of points among only those paths that connect nearest neighbors using a technique such as Floyd's algorithm.

Step3. Apply classical MDS to embed the data in a d -dimensional Euclidean space so as to preserve these geodesic distances.

3. Nonlinear Dimension Reduction of Vowel Signal Data

The dimension reduction method based on manifold learning focuses on preserving some geometric attributes in manifold after dimension reduction. Some scholars have studied the geometric structure of natural speech data by using the acoustic pipe model generated by speech, and verified the correctness of manifold hypothesis in speech data [7].

In this study, the vowel signal is analyzed based on data dimension reduction, and the signal processing tools in MATLAB software and Voicebox toolbox are used [11]. The experimental data are monophthongs [e] and [u] with a sampling frequency of 16 kHz. Firstly, the signal is preprocessed by framing and windowing. Several speech sampling points are divided into a frame (denoted by N), in which the signal characteristics are considered to remain unchanged [12]. In order to make the transition between frames smooth, there will be partial overlap between frames (as shown in Fig. 1), and the overlapping area is the frame shift (denoted by M), and the ratio of frame shift to frame length is generally $0 \sim 1/2$. In order to make both ends of the signal edge continuous after framing, in practical problems, each frame signal should be multiplied by a window function to prevent truncation effect and spectrum leakage [13]. Therefore, in this paper, firstly, the Hamming window with frame length of 25 ms and frame shift of 16ms is used to preprocess the vowel signal, and the simulated non-stationary speech signal is changed into a short-time stationary signal [1]. Then, two nonlinear dimension reduction algorithms, ISOMAP and LLE, are used to reduce the dimension of the high-dimensional vowel signal, and the two-dimensional geometric structure of the vowel signal is obtained for subsequent analysis while keeping the manifold structure of the original high-dimensional signal unchanged as much as possible.

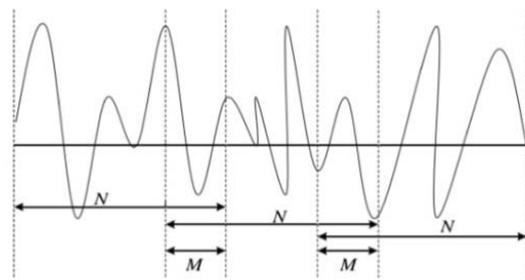


Fig. 1. Schematic diagram of frame length and frame movement

Both LLE and ISOMAP algorithms need to manually adjust the nearest neighbor parameter K in order to obtain the best experimental results. When k is 9, 12, 15, 18, 21, 23, 25, and 28, the preprocessed experimental data are subjected to ISOMAP and LLE dimension reduction processing, and the LLE and ISOMAP two-dimensional geometric structures of vowel signals [e] and [u] are obtained respectively, as shown in the following Fig.2, Fig.3 and Fig.4. Experiments on vowel signals show that there is a low-dimensional manifold structure in vowel signal data, that is, the high-dimensional vowel signals are nonlinear manifolds. Comparing the LLE and ISOMAP two-dimensional embedding results of vowel signals given below, it can be seen that the low-dimensional embedding results of ISOMAP change little and have certain regularity when the K value changes little; On the contrary, the low-dimensional geometric structure obtained by LLE is distorted seriously when the value of k changes little, so it can be considered that the two-dimensional embedded graph obtained by ISOMAP dimension reduction can express the essential characteristics of vowel signals more stably. Experiments on more vowel signals show that the two-dimensional ISOMAP embedding results obtained from different vowel signals are different and have their own characteristics. Compared with LLE low-dimensional embedding results, the low-dimensional ISOMAP embedding results are more stable, so it can be considered to be applied to subsequent classification analysis.

The generation of speech is a very complicated process. How to extract and select the characteristic parameters that can represent the essential characteristics of speech signals is an important link in speech recognition and other technologies. The data dimension reduction technology based on manifold learning can not only get the low-dimensional embedding results of signals [12], but also get LLE and ISOMAP feature parameters representing vowel signal features, which is expected to be applied to speech recognition and other technologies in order to discover more signal features.

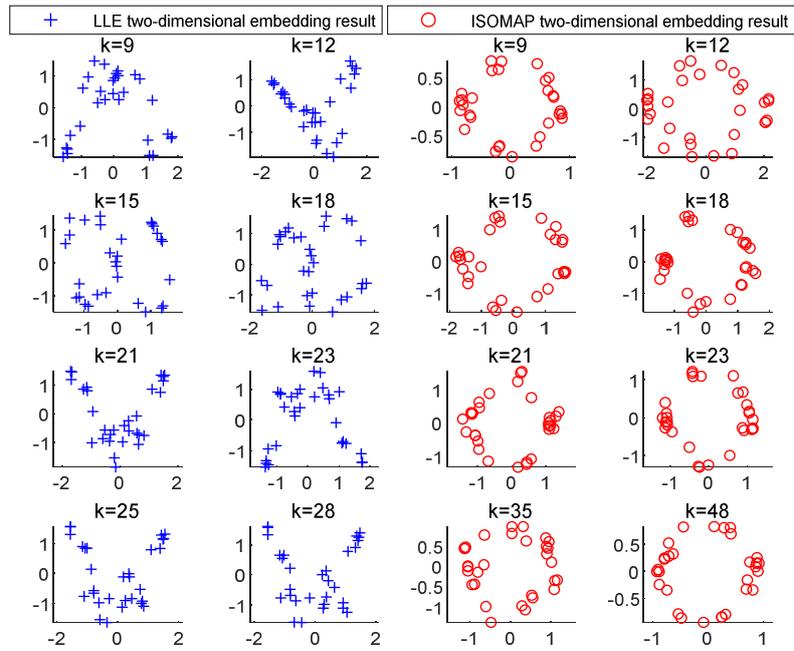


Fig. 2. Two-dimensional embedding results of LLE and ISOMAP under different k values of vowel [e]

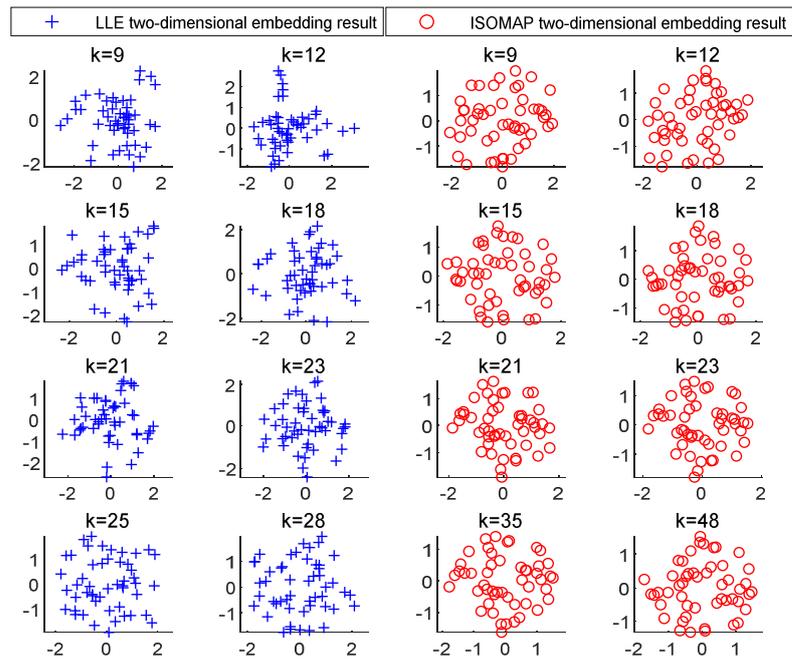


Fig. 3. Two-dimensional embedding results of LLE and ISOMAP under different k values of vowel [o]

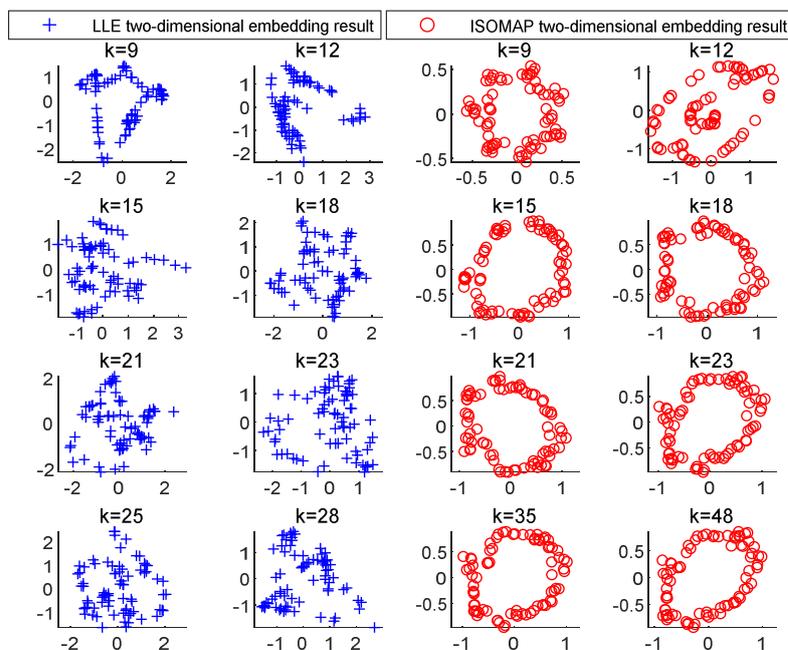


Fig. 4. Two-dimensional embedding results of LLE and ISOMAP under different k values of vowel [u]

4. Conclusion

In order to solve the problem that the traditional feature analysis method based on linear system theory can't find the essential features of speech signals well, we studied data dimension reduction method based on manifold learning and applies it to the analysis of vowel signals. The simulated speech signal is preprocessed, and the preprocessed signal is reduced in dimension by using ISOMAP and LLE. By manually adjusting the nearest neighbor parameter K , the 2D embedding result of high-dimensional speech signal is obtained. Experiments show that the dimension reduction method based on manifold learning can eliminate redundant information and realize two-dimensional visualization of signals while preserving the high-dimensional manifold structure of original signals as much as possible. However, there are some subjective factors in choosing the nearest neighbor parameter K , which has a great influence on the analysis results. The way to choose the appropriate parameter K is the key content future research.

References

- [1] Claudio Turchetti, Laura Falaschetti. A manifold learning approach to dimensionality reduction for modeling data. *Information Sciences*, 2019, 491, 16-29.
- [2] Enzhi Gao. Research on dimension reduction technology based on manifold learning and its application. Jiangnan University, 2012(14)
- [3] Yugong Xu. Speech feature extraction based on manifold learning. Shandong University, 2017 (11)
- [4] Zhong Zhao, Guocan Feng, Jiehua Zhu, Qi Shen. Manifold learning: dimensionality reduction and high dimensional data reconstruction via dictionary learning. *Neurocomputing*, 2016, 216, 268-285.
- [5] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*. 2000, 290(5500): 2319-2323.
- [6] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*. 2000, 290 (5500): 2323-2326.
- [7] Hui Liu, Jun 'an Yang, Yi Wang. Research on feature extraction method of acoustic target based on manifold learning. *Journal of Information Physics*, 2011, 60 (7): 1729-1736.
- [8] Shaeela Ayesha, Muhammad Kashif Hanif, Ramzan Talib. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 2020, 59, 44-58.
- [9] An investigation of manifold learning for Chinese handwriting analysis. *International Journal of Systems and Control*, 2007, (03): 213-220.
- [10] Yufeng Xiao. Research and DSP implementation of speech feature extraction based on ISOMAP. Hunan University, 2013(07)
- [11] Paul S Bibin Sam, Glittas Antony Xavier, Gopalakrishnan Lakshminarayanan. A low latency modular level deeply integrated MFCC feature extraction architecture for speech recognition. *Integration*, 2021, 76, 69-75.
- [12] Garima Sharma, Kartikeyan Umopathy, Sridhar Krishnan. Trends in audio signal feature extraction methods[J]. *Applied Acoustics*, 2020, 158.
- [13] M. M. Hasan, H. Ali, M. F. Hossain and S. Abujar, "Preprocessing of continuous bengali speech for feature extraction," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-4.

**Creative Commons Attribution License 4.0
 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US